

FIGURE 1A

Sequence of cadherin 3 (GenBank accession number NP_001784)

SEQ ID NO:1

MGLPRGPLASLLLLQVCWLQCAASEPCRAVFREA EVTLEAGGAEQEPGQALGK
VFMGCPGQEPALFSTDNDDFTVRNGETVQERRSLKERNPLKIFPSKRILRRHKRD
WVVA PISVPENGKGPFQRLNQLKSNKDRDTKIFY SITGPGADSPPEGVFAVEKE
TGWLLL NKPLDREEIAKYELFGH AVSENGASVEDPMNISIIVTDQNDHKPKFTQD
TFRGSVLEGVLPGTSVMQVTATDEDDAIYTYNGVVAYSISQEPKDPHDLMTI
HRSTGTISVISSGLDREKVPEYTLTIQATDMDGDGSTTTAVAVVEILDANDNAPM
FDPQKYEAHVPENAVGHEVQRLTVTDLDAPNSPAWRATYLMGGDDGDHFTITT
HPESNQGILTTRKGLDFEAKNQHTLYVEVTNEAPFVLKLPTSTATIVVHVEDVNE
APV FVPPSKVVEVQEGIPTGEPVCVYTAEDPDKENQKISYRILRDPAGWLAMDPD
SGQVTAVGTL DREDEQFVRNNIYEVMLV LAMDN GSPPTTGTGTL LTLIDVNDHG
PVPEPRQITICNQSPVRHVLNITDKDLSPTSPFQAQLTDDSDIYWTAEVNEEGDT
VVL S LKKFLKQDTYDVHLSLSDHGNKEQLTVIRATVCDCHGHVETCPGPWKGG
FILPVLGAVLALLFLLLVL LLLVRKKRKIKEPLLLPEDDTRDNV FYYGEEGGGEE
DQDYDITQLHRGLEARPEVVL RNDVAPT IPTPMYRPRPANPDEIGNFIIENLKAA
NTDPTAPPYDTLLVFDYEGSGSDAASLSSLTSSASDQDQDYDYLNEWGSRFKKL
ADMYGGGEDD

FIGURE 1B

Sequence of matrix metalloproteinase 14 (GenBank accession number NP_004986)

SEQ ID NO:2

MSPAPRPPRCLLLPLLTLGTALASLGSAQSSSFSP EAWLQQYGYLPPGDLRHTHTQ
RSPQSL SAAIAAMQKFYGLQVTGKADADTMKAMRRPRCGVPDKFGAEIKANVR
RKRYAIQGLKWQHNEITFCIQNYTPKVGEYATYEAIRKA FRVWESATPLRFREVP
YAYIREGHEKQADIMIFFAEGFHGDSTPFDGEGGFLAHAYFPGPNIGGDTHFDSA
EPWTVRNEDLNGNDIFLVAVHELGHALGLEHSSDPSAIMAPFYQWMDTENFVLP
DDDRRG IQQLYGGESGFPTKMPPQPR TTSRPSVPDKPKNPTYGPNICDGNFDTVA
MLRGEMFVFKERWFWVRNNQVMDGYPMPIGQFWRGLPASINTAYERKD GKF
VFFKGDKHWVFDEASLEPGYPKHIKELGRGLPTDKIDAALFWMPNGKTYFFRGN
KYYRFNEELRAVDSEYPKNIKVWEGIPESPRGSFMGSDEVFTYFYKGNKYWKFN
NQKLKVEPGYPKSALRDWMGCPSGGRPDEGTEEE TEVIIIEVDEEGGGAVSAAA
VVLPLV LLLLVLAVGLAVFFFRRHGTPRRLLYCQRSLLDKV

FIGURE 1C

(1)

Sequence of cadherin EGF LAG seven-pass G-type receptor 2 (GenBank accession number NP_001399)

SEQ ID NO:3

MRSPATGVPLPTPPPLLLLLLLLLPPPLLGDQVGPCRSLGSRGRGSSGACAPMG
WLCPSASNLWLYTSRCRDAGTELTGHLVPHHDGLRVWCPSEAHIPPAPEG
CPWSCRLLGIGGHLSPQGKLTLP EEHPCLKAPRLRCQSKLAQAPGLRAGERSPE
ESLGGRKRNVNTAPQFQPPSYQATVPENQPAGTPVASLRAIDPDEGEAGRLEYT
MDALFDSRSNQFFSLDPVTGAVTTAEELDRETKSTHVFRVTAQDHGMPRRSALA
TLTILVTDNDHDPVFEQQEYKESLRENLEVG YEVLTVRATDGDAPPNANILYRL
LEGSGGSPSEVFEIDPRSGVIRTRGPVDREEVESYQLTVEASDQGRDPGPRSTTAA
VFLSVEDDNDNAPQFSEKRYVVQVREDVTPGAPVLRVTASDRDKGSNAV VHYSI
MSGNARGQFYLDAQTGALDVVSPLDYETTKEYTLRVRAQDGRPPLSNVSGLV
TVQVLDINDNAPIFVSTPFQATVLESVPLGYLVLVHVQAIDADAGDNARLEYRLAG
VGHDFPFTINNGTGWISVAAELDREEVDFYSFGVEARDHGTPALTASASVSVTVL
DVNDNNPTFTQPEYTVRLNEDAAVGTSVTVSAVDRDAHSVITYQITSGNTRNR
FSITSQSGGGLVSLALPLDYKLERQYVLA VTASDGTRQDTAQIVNVVTDANTHRP
VFQSSHYTVNVNEDRPAGTTVVLISATDEDTGENARITYFMEDSIPQFRIDADTG
AVTTQAELDYEDQVSYTLAITARDNGIPQKSDTTYLEILVNDVNDNAPQFLRDSY
QGSVYEDVPPFTSVLQISATDRDSGLNGRVFYTFQGGDDGDGDFIVESTSGIVRT
LRRLDRENV AQYVLRAYAVDKGMPPARTPMEVTVTVLDVNDNPPVFEQDEFDV
FVEENSPIGLAVARVTATDPDEGTNAQIMYQIVEGNIPEVFQLDIFSGELTALVDL
DYEDRPEYVLVIQATSAPLVSRATVHVRLDDRNDNPPVLGNFEILFN NYVTNRSS
SFPGGAIGRVP AHDPDISDSLTSFERGNELSLVLLNASTGELKLSRALDNNRPLE
AIMSVLVSDGVHSVTAQCALRVTIITDEMLTHSITLRLEDMSPERFLSPLLGLFIQA
VAATLATPPDHVVVFNVQRDTDAPGGHILNVSLSVGQPPGPGGGPPFLPSEDLQE
RLYLNRSLT AISAQRVLPFDDNICLREPCENYMRCVSVLRFDSSAPFIASSSVLFR
PIHPVGGLRCRCPPGFTGDYCETEVDLCYSRPCGPHGRCRSREGGYTCLCRDGYT
GEHCEVSARSGRCTPGVCKNGGTCVNLLVGGFKCDCPSGDFEKP YCQVTTRSFP
AHSFITFRGLRQRFHFTLALS FATKERDGLLYNGRFNEKHDFVALEVIQEQVQL
TFSAGESTTTVSPFVPGGVSDGQWHTVQLKYYNKPLL GQTGLPQGPSEQKVAVV
TVDGC DTGVALRFGSVLGNYSCAAQGTQGGSKKSLDLTGPLLLGGVPDLPESFP
VRMRQFVGCMRNLQVDSRHIDMADFIANNGTVP GCPAKKNVCDSNTCHNGGT
CVNQWDAFSCCPLGFGGKSCAQEMANPQHFLGSSLVAWHGLSLPISQPWYLSL
MFRTRQADGVLLQAITRGRSTITLQLREGHVMLSVEGTGLQASSLRLEPGRAND
GDWHHAQLALGASGGPGHAILSFDY GQQRAEAGNLGPRLHGLHLSNITVGGIPGP
AGGVARGFRGCLQGVRVSDTPEGVNSLDP SHGESINVEQGC SLPDPCDSNPCPA
NSYCSNDWDSYSCSDPGYYGDNCTNVCDLNPCEHQSVCTRKPSAPHGYTCEC
PPNYLG PYCETRIDQPCPRGWWGHPTCGPCNCDVSKGFDPCNKTSGECHCKEN
HYRPPGSPTCLLCDCYPTGSLSRVCDPEDGQCPCPKPGVIGRQC DRCDNPFAEVT
NGCEVNYDSCPRAIEAGIWWPRTRFGLPAAAPCPKGSFGTAVRHCDEHRGWLPP
NLFNCT SITFSELKGFAERLQRNESGLDSGRSQQ LALLRNATQHTAGYFGSDVK
VAYQLATRLLAHESTQRGFGLSATQDVHFTENLLRVGSALLDTANKRHWELIQQ

FIGURE 1C

TEGGTAWLLQHYEAYASALAQNMRYHTYLSPTIVTPNIVISVVRLDKGNFAGAK
LPRYEALRGEQPPDLETTVILPESVFRETTPPVVRPAGPGEAQEPEELARRQRRHPE
LSQGEAVASVIIYRTLGLLPHNYDPDKRSLRVPKRPIINTPVVSISVHDDEELLPR
ALDKPVTVQFRLLTEERTKPICVFWNHSILVSGTGGWSARGCEVVFRNESHVSC
QCNHMTSFAVLMDVSRRENGEILPLKTLTYVALGVTLAALLLTFFFLTLRLRS
NQHGIRRNLTAAALGLAQLVFLGINQADLPFACTVIAILLHFLYLCTFSWALLEAL
HLYRALTEVRDVNTGPMRFYYMLGWGVPAFITGLAVGLDPEGYGNPDFCWLSI
YDTLIWSFAGPVAFVSMVFLYILAAARASCAAQRQGFEEKKGPVSGLQPSFAVLL
LLSATWLLALLSVNSDTLLFHLYLFATCNCIQGPFIFLSYVVLSKEVRKALKLACSR
KPSDPALTTKSTLTSSYNCPSPYADGRLYQPYGDSAGSLHSTSRSGKSQPSYIPF
LLREESALNPGQGPPGLGDPGSLFLEGQDQQHDPDTSDDSDLSLEDDQSGSYAST
HSSDSEEEEEEEEEAAFPGEQGWDSLLGPGAERLPLHSTPKDGGPGPGKAPWPG
DFGTTAKESSGNGAPEERLRENGDALSGSLGPLPGSSAQPHKGILKKKCLPTIS
EKSSLLRLPLEQCTGSSRGSSASEGSRGGPPRPPPRQSLQEQLNGVMPIAMSIKA
GTVDEDSSGSEFLFFNFLH

bioRxiv preprint doi: <https://doi.org/10.1101/072601>; this version posted October 1, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1D

Peptides for antibodies that bind to cadherin3 (GenBank accession number NP_001784):

RAVFREA EVTLEAGGAEQE (SEQ ID NO:4)

QEPALFSTDNDDFTVRN (SEQ ID NO:5)

QKYEAHVPENAVGHE (SEQ ID NO:6)

Peptides for antibodies that bind to matrix metalloproteinase 14 (GenBank accession number NP_004986):

AYIREGHEKQADIMIFFAE (SEQ ID NO:7)

DEASLEPGYPKHIKELGR (SEQ ID NO:8)

RGSFMGSDEVFTYFYK (SEQ ID NO:9)

Peptides for antibodies that bind to anti-cadherin EGF LAG seven-pass G-type receptor 2 (GenBank accession number NP_001399):

QASSLRLEPGRANDGDWH (SEQ ID NO:10)

ELKGFAERLQRNESGLDSGR (SEQ ID NO:11)

RSGKSQPSYIPFLLREE (SEQ ID NO:12)

Peptides for antibodies that bind to anti-cytokeratin17:

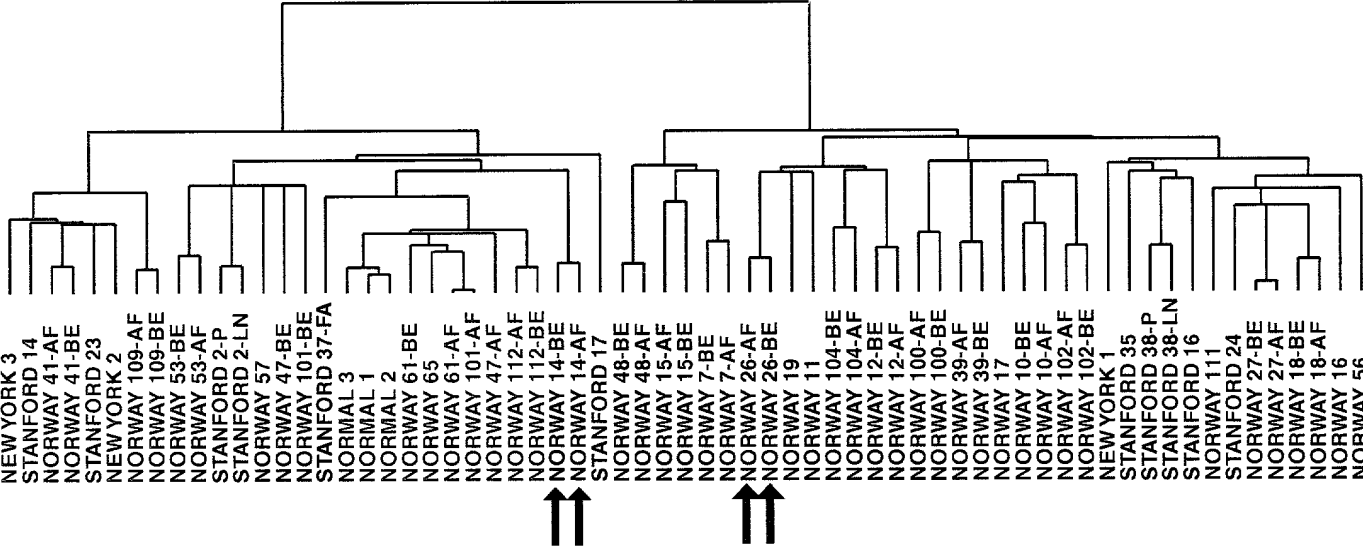
KKEPVTTRQVRTIVEE (SEQ ID NO:13)

QDGKVISSREQVHQTTR (SEQ ID NO:14)

SSSIKGSSGLGGGSS (SEQ ID NO:15)

FIGURE 2

Intrinsic Gene Subset



Epithelial-Enriched Gene Subset

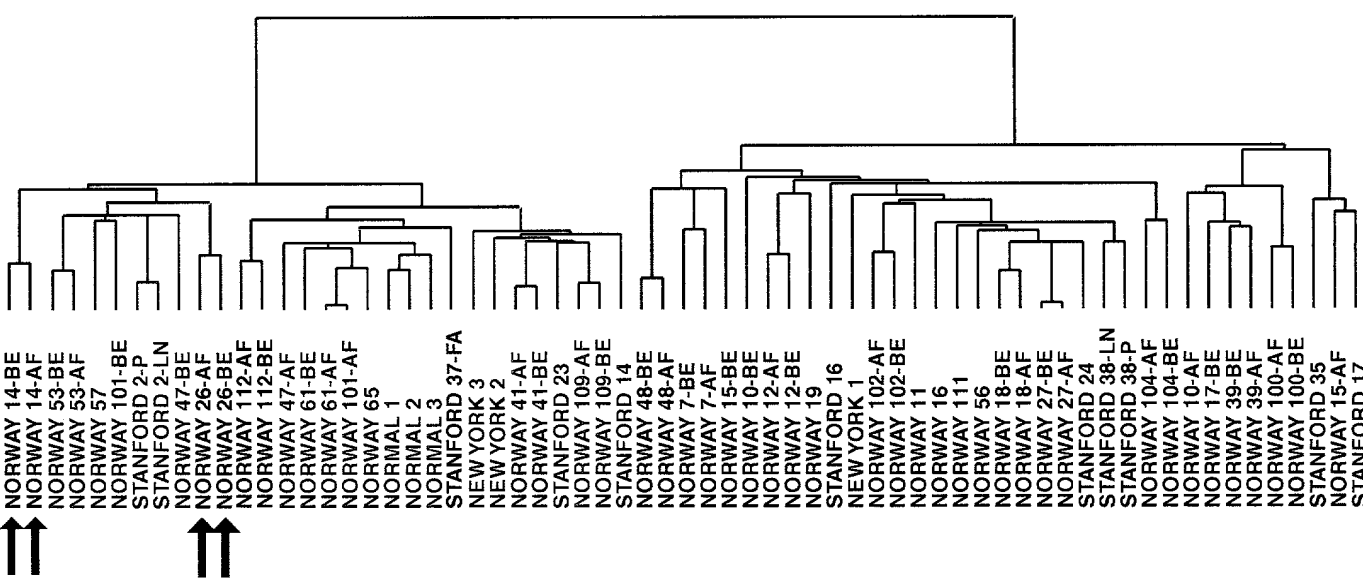
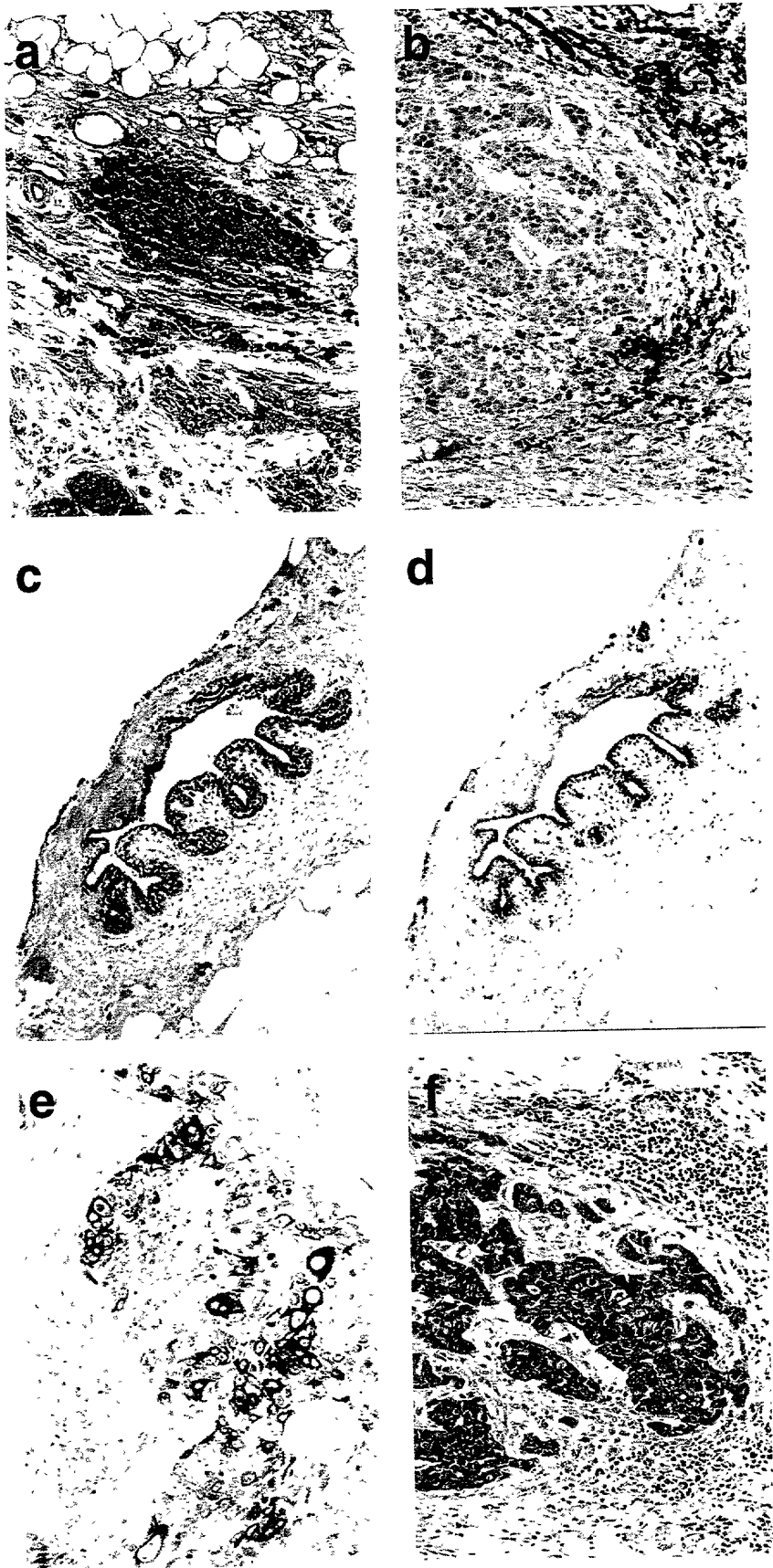


Figure 3



09916849.072601
T09220" 6489T660

S0158

1591649 07004

188

62

49

38

28

18

1:100

FIGURE 4A

S0144

1:200

1:500

1:1000

FIGURE 4B

S0137

1:200

1:500

1:1000

FIGURE 4C

Figure 5A

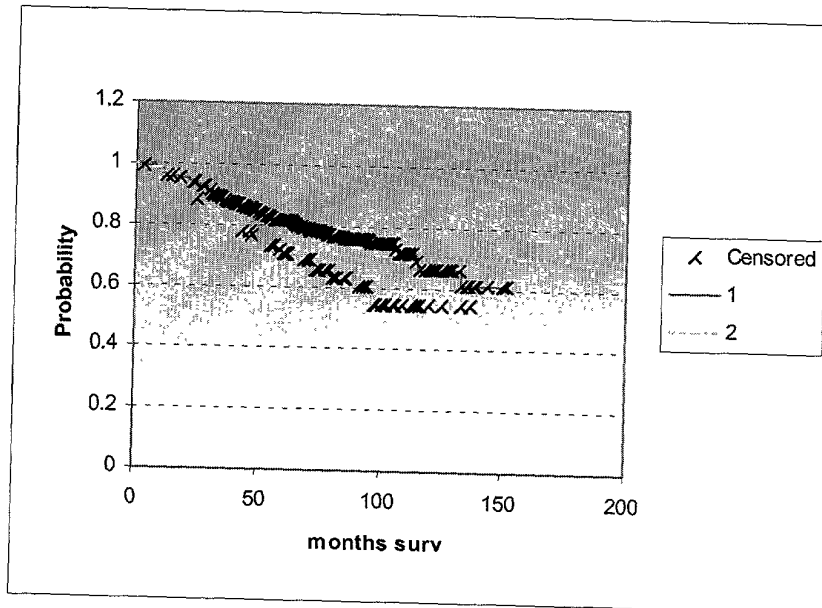
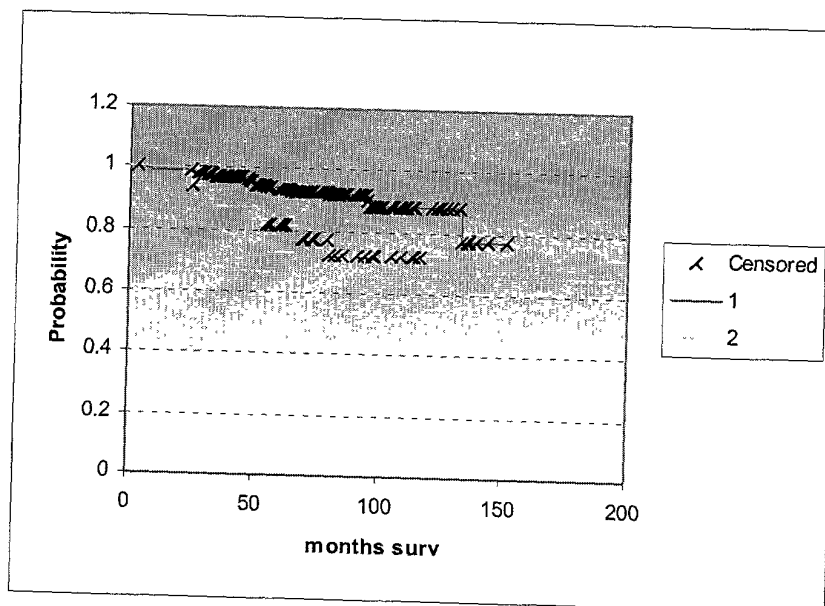
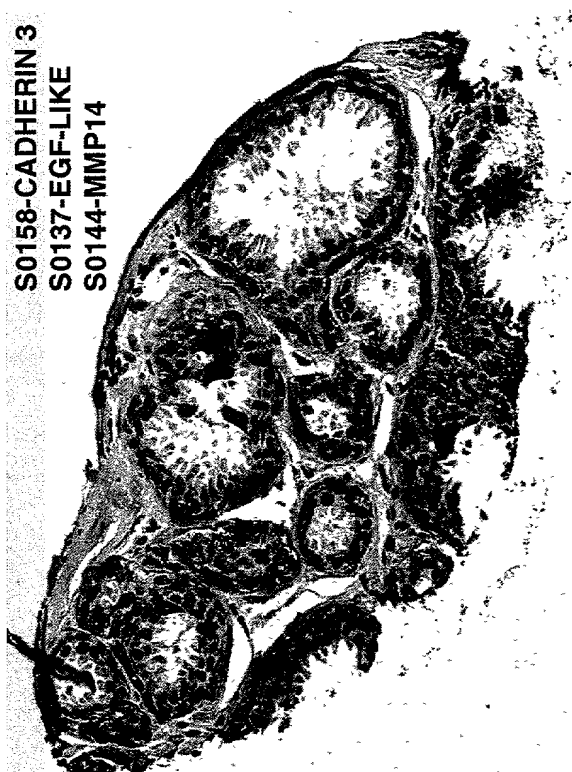


Figure 5B

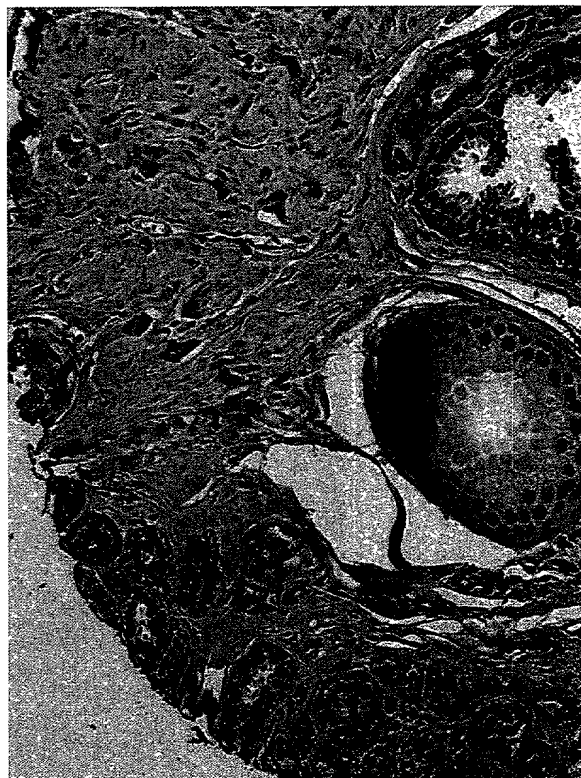


A. ck5/6



S0158-CADHERIN 3
S0137-EGF-LIKE
S0144-MMP14

B. s0158



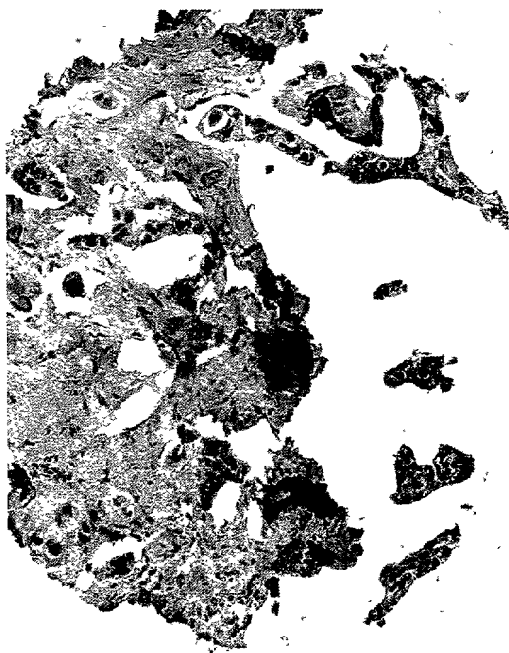
C. s0137



D. s0144

FIGURE 6

A. CK5/6



B. S0137



C. S0158

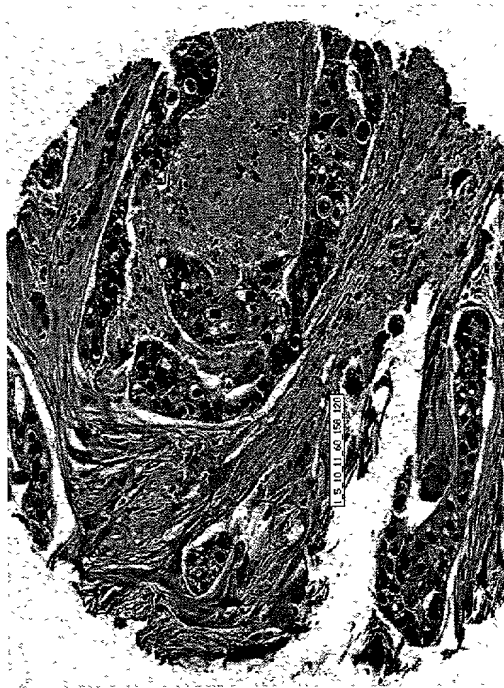


FIGURE 7